DOCUMENT RESUME

ED 409 381 TM 026 967

AUTHOR	Stanley, Julian C.; Stumpf, Heinrich
TITLE	Gender Differences, Especially on Fifty College Board Achievement Tests.
PUB DATE	Apr 97
NOTE	25p.; Paper presented at the Annual Meeting of the Eastern Psychological Association (Washington, DC, April 1997).
PUB TYPE	Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Achievement Tests; College Entrance Examinations; Correlation; *Course Selection (Students); Followup Studies; Higher Education; *Scores; *Sex Differences; *Test Results
IDENTIFIERS	*Advanced Placement Examinations (CEEB); *College Board Achievement Tests; College Entrance Examination Board

ABSTRACT

In a follow-up to findings published by H. Stumpf and J. Stanley (1996), the gender-related differences in enrollment in and scores on the College Board Achievement (SAT II) and Advanced Placement (AP) tests were studied. Differences in scores turned out to be rather stable from 1982 (for the Achievement tests) and 1984 (for the AP tests) through 1996, with 12 of the 21 SAT II tests favoring males and 2 favoring females. The differences in scores on the Achievement Test in American History and the AP Computer Science A and AB examinations, however, declined considerably in the period studied here. While there were substantial gains in the numbers of females scoring high on the Physics and Mathematics II Achievement tests, the low enrollment of female students in AP Computer Science A and AB continued to be a matter of concern. As found previously, there was a strong correlation between the percentages of males taking the two sets of tests and the gender-related differences in scores on them. (Contains 3 tables, 3 figures, and 19 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

PERMISSION TO REPRODUCE AND
DISSEMINATE THIS MATERIAL
HAS BEEN GRANTED BY
Julian Stanley

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED 409 381

Gender Differences, Especially on Fifty College Board Achievement Tests*

Julian C. Stanley and Heinrich Stumpf

Johns Hopkins University

TM026967

BEST COPY AVAILABLE

Abstract

In a follow-up on findings published by Stumpf and Stanley (1996), we examined gender-related differences in enrollment in and scores on the College Board Achievement (SAT II) and Advanced Placement (AP) tests. Differences in scores turned out to be rather stable from 1982 (for the Achievement tests) and 1984 (for the AP tests) through 1996, with 12 of the 21 SAT II tests favoring males and 2 favoring females, and 18 of the AP examinations favoring males and 6 favoring females. The differences in scores on the Achievement test in American History and the AP Computer Science A and AB examinations, however, declined considerably in the period studied here. While there were substantial gains in the numbers of females scoring high on the Physics and Mathematics II Achievement tests, the low enrollment of female students in AP Computer Science A and AB continued to be a matter of concern. As found previously, there was a strong correlation between the percentages of males taking the two sets of tests and the gender-related differences in scores on them.

Gender Differences, Especially on Fifty College Board Achievement Tests

In this presentation we are concerned mainly with achievement tests designed for selection into college or placement therein. Scores on such tests can affect the lives of examinees much more than the usual achievement tests in high school do. Of course, we are dealing with highly self-selected test takers. Therefore, we make no claim or effort to generalize to gender differences among representative or randomly chosen samples of boys and girls. Also, we present the "whats," rather than saying much about the "whys," which is the topic of other presentations today.

This report is the most recent outcome of twenty-five years of studies at Johns Hopkins University of gender differences on cognitive tests. They began in March of 1972 at the first talent search conducted by Stanley's Study of Mathematically Precocious Youth, SMPY (Keating & Stanley, 1972; Stanley, 1973; Stanley, Keating, & Fox, 1974). In 1980, our research gained notoriety as the result of lurid press coverage of a brief report in the professional journal *Science* (see Benbow & Stanley, 1980, 1981, 1982, 1983).

More recently, my colleagues and I have produced a number of relevant articles (Stanley et al., 1992; Stanley, 1994; Stanley, Stumpf, & Cohn, in press; Stumpf & Stanley, 1996, 1997). They involve far more than a hundred tests, the majority of which were constructed by the Educational Testing Service. Seven

generalizations from these studies are:

1) The largest gender differences favoring males occur for theoretical evaluative attitude and mechanical reasoning, about a standard deviation each. Relatively large differences were also found in the area of spatial ability, especially for mental rotation (Stumpf, 1993; Stumpf & Eliot, 1995).

2) The largest gender differences favoring females occur for aesthetic and social service evaluative attitudes, spelling in the twelfth grade (about half a standard deviation), language usage, and clerical speed and accuracy. Differences in favor of females of about half a standard deviation in size were also found for memory performance (Stumpf & Jackson, 1994; Stumpf & Eliot, 1995).

3) The Medical College Admissions Test (MCAT) shows about the same pattern of gender differences favoring males that college entrance and placement tests do: least on reading, most on physics. The Law School Admissions Test (LSAT) showed no appreciable gender differences.

4) On all 17 Graduate Record Examination subject tests males averaged higher scores than females, from one-sixth of a standard deviation for Psychology to more than three-fourths of a standard deviation for Political Science.

5) Among intellectually bright students, substantial differences occur in elementary school. Also, Robinson and her associates (1996) found some among preschoolers.

6) Despite the fact that there are mean differences between

the sexes on a number of tests, the factor structures of them are highly similar for males and females (Stumpf & Jackson, 1994; Stumpf & Eliot, 1995).

7) The gender differences from most of the 50 tests we studied for this presentation seemed to have remained fairly constant from 1982 or 1984 through 1996. Exceptions in our analyses (Stumpf & Stanley, 1996) are only the College Board high school Achievement test in American History, the Advanced Placement Program (AP) test in Computer Science, and the Cube Perspectives Test of spatial ability (Stumpf & Klieme, 1989), all of which favor males now less than then.

Figure 1 shows the trends for American History during the fifteen-year period. The *d*-values ("effect sizes," which are

---

Insert Figure 1 about here.

---

standardized differences between means) dropped from forty-three hundredths of a standard deviation in 1982 to only twenty-three hundredths of a standard deviation in 1996. The upper-tail ratio (cf. Feingold, 1995), which is the percentage of males scoring 700 or more divided by the percentage of females scoring 700 or more, dropped from 2.78 to 1.46. The lower-tail ratio, the percentage of females scoring less than 300 divided by the percentage of males scoring less than 300, dropped more erratically from 2.87 to 1.56.

The curves for the two AP Computer Science tests are

similar to those for American History. See Figure 2, where

Insert Figure 2 about here.

---

results for Level AB (two semesters of college credit) are plotted. The *d*-values plummeted from fifty-nine-hundreds of a standard deviation in 1984, the first year the test was administered, to only sixteen-hundredths of a standard deviation in 1996. The upper-tail ratio, percentage of males who score 5 (the highest possible) divided by percentage of females scoring 5, follows the same trend. The lower-tail ratio, percentage of females scoring 1 (the lowest possible) divided by percentage of males scoring 1, declines less sharply.

Results for the Computer Science Level A test (one semester of college credit) are available for only the six years, 1991 through 1996, that it has been offered. As Figure 3 shows, the *d*-values dropped from .57 to .33. The ratios followed suit.

Insert Figure 3 about here.

On no other College Board Achievement or AP test did we find systematic evidence of declines in gender differences. Of course, any declines or increases are difficult to interpret. The type of students taking a test may change from year to year. Committees constructing the tests change, and so probably do the test specifications. Recently, too, ETS has been studying items

for unusual gender differences and replacing some. This seems likely to result in more declines, so henceforth no one may be able to assess "real" changes in achievement.

This raises an interesting issue. For example, if the specifications used in constructing an achievement test call for items concerning Napoleon's defeat in the Waterloo Campaign, and girls score poorly on them, should ETS substitute items concerning Napoleon's family life? What would be the logic justifying such a switch?

On a broader front, let's turn to the year 1996 results for the 21 College Board Achievement tests. They are shown in Table 1. Focus on columns 4 and 5, the effect size (*d*) and the UTR

Insert Table 1 about here.

(upper-tail ratio). On eight tests both of these systematically favor males. On no tests do both favor females. Overall, males have a slight *d*-advantage (.21) and a 1.71-to-1 lead on the upper-tail ratio.

The largest effect size, half a standard deviation, is for Physics. How big is that discrepancy? If the scores are normally distributed and females score at the middle of the distribution, that is, at the 50th percentile, males would score at the 69th percentile.

The smallest effect size tabulated is one-fifth of a standard deviation. This corresponds to the 50th percentile of

females versus the 58th percentile of males.

Upper-tail ratios can be interpreted more straightforwardly. They range from 2.40 to 1 for World History to 1.27 to 1 for German with Listening, the only test favoring females at the top or the middle. Thus, percentage-wise, males had a one and four-tenths advantage on World History, whereas females had a bit more than a one-fourth advantage on German with Listening.

Please note that the nine tests not favoring males all heavily involve language usage. Even then, Latin, French, Modern Hebrew, and Italian favored males.

Some have suggested that gender differences would be lessened by replacing SAT-V and SAT-M with achievement tests. Probably not. If colleges required one mathematics Achievement test, one science test, and one English test, females would still have two strikes against themselves and no test on which they significantly excel males. Requiring six Achievement tests --say, adding history, a foreign language, and one elective-- wouldn't seem to help much, especially for applicants to selective colleges. Also, the cost of taking achievement tests would undoubtedly greatly exceed the cost of taking SAT-V and SAT-M.

We now turn to the 29 College Board Advanced Placement

Insert Table 2 about here.

Program (AP) tests for 1996 (see Table 2). In columns 4, 5, and 6 of the table are shown effect sizes, upper-tail ratios, and lower-tail ratios. On 13 tests, all three of these difference indices uniformly favor males, versus two for females. Eighteen of the upper-tail ratios favor males, versus the two for females. At the bottom of the distribution, 15 lower-tail ratios favor males, versus six for females. Overall, however, males excel only on the upper-tail ratio, and just moderately (1.36) even there.

It appears that the AP tests, nearly all of which consist of half multiple-choice items and half open-ended ("essay") items, are a little kinder to females than the Achievement tests are. It is well known that, relative to boys, girls tend to perform better on open-ended questions than on multiple-choice items, especially when language usage skills are appreciably involved (see further comments about this in Stumpf & Stanley, 1996).

For both sets of achievement tests, the three gender-differences statistics correlate negatively with the percentage of females who take the test: the fewer who take a given test, the worse they tend to do relative to the males taking the test; the more who take it, the better they tend to do. Correlation

---

Insert Table 3 about here.

---

coefficients with the percentages of AP test takers who are

female are -.71 for effect size, -.62 for UTR, and -.66 for LTR. The figures for the 21 Achievement tests are lower: -.64, -.28, and -.10 (the latter one for an N of only 10), respectively (see also Table 3).

We conclude that, for whatever reasons, males who take the College Board Achievement and AP tests are appreciably advantaged relative to female takers, both with respect to their scores and also to the percentage who sign up for some of the most important tests. Low points of the latter are that only 27 percent of examinees who took the Physics Achievement test in 1996 were female, and only 12 percent of the AP two-semester Computer Science test takers were female. Of course, much research about causes is needed.

We can close on a brighter note, however (Stanley, in press). Urging young women to take the Physics and the Mathematics Level II (precalculus) Achievement tests has paid off well from 1982, when our study began, to 1996. Initially, only 200 females had scored 700-800 on Physics, whereas 15 years later 877 had. That's a phenomenal 338 percent increase! For Math II the respective figures are 3429 and 6329, an increase of 85 percent.

There is something strange about that 85 percent math gain, however: two years earlier it was far greater, 3429 versus 9032, an increase of 163 percent. Apparently the transition from Math II without calculator to Math II with calculator eliminated many high-scoring females, even though enrollment numbers did not

decline (31,270 in 1994, and 33,264 in 1996). Surely, ETS must have noticed this catastrophic drop. Is it, along with the poor representation of females on the AP Computer Science tests, an indication of dislike or ineptitude for things mechanical? If so, how does one account for the huge increase in high scores by females on the Physics Achievement test? Perhaps girls merely need more experience with calculators and computers.

A part of the decrease, however, could also be due to recentering of the SAT scores in April 1995. SAT Verbal and Mathematical scores of examinees who take a certain Achievement test form the basis for the location of the test's scores on the 200-to-880 College Board standard scale.

In any event, this considerable worsening of females' accomplishment on the Math II Achievement test in a two-year period needs investigating.

References

Benbow, C. P., & Stanley, J. C. (1980). Sex differences in mathematical ability: Fact or artifact? *Science, 210*, 1262-1264.

Benbow, C. P., & Stanley, J. C. (1981). Mathematical ability: Is sex a factor? *Science, 212*, 118 & 121.

Benbow, C. P., & Stanley, J. C. (1982). Consequences in high school and college of sex differences in mathematical reasoning ability: A longitudinal perspective. *American Educational Research Journal, 19*, 598-622.

Benbow, C. P., & Stanley, J. C. (1983). Sex differences in mathematical reasoning ability: More facts. *Science, 222*, 1029-1031.

Feingold, A. (1995). The additive effects of differences in central tendency and variability are important in comparisons between groups. *American Psychologist, 50*, 5-13.

Keating, D. P., & Stanley, J. C. (1972). Extreme measures for the exceptionally gifted in mathematics and science. *Educational Researcher, 1*, 3-7.

Robinson, N. M., Abbott, K. D., Berninger, V. M., & Busse, J. (1996). The structure of abilities in math-precocious young children: Gender similarities and differences. *Journal of Educational Psychology, 88*, 341-352.

Stanley, J. C. (1973). Accelerating the educational progress of intellectually gifted youths. *Educational Psychologist, 10*, 133-146.

Stanley, J. C. (1994). Gender differences for able elementary school students on above-grade-level achievement tests. In N. Colangelo, S. G. Assouline, & D. L. Ambroson (Eds.), *Talent development*, vol. 2 (pp. 141-148). Dayton, Ohio: Ohio Psychology Press.

Stanley, J. C. (in press). Varieties of intellectual talent. *Journal of Creative Behavior*.

Stanley, J. C., Benbow, C. P., Brody, L. E., Dauber, S., & Lupkowski, A. E. (1992). Gender differences on eighty-six nationally standardized aptitude and achievement tests. In N. Colangelo, S. G. Assouline, & D. L. Ambroson (Eds.), *Talent development*, vol. 1 (pp. 42-65). Unionville, NY: Trillium Press.

Stanley, J. C., Keating, D. P., & Fox, L. H. (1974). *Mathematical talent: Discovery, description, and development*. Baltimore, MD: Johns Hopkins University Press.

Stanley, J. C., Stumpf, H., & Cohn, S. J. (in press). Ipsative evaluative attitudes versus cognitive abilities and vocational interests of bright male versus female seventh graders. In N. Colangelo, S. G. Assouline, & D. L. Ambroson (Eds.), *Talent Development*, vol 3. Dayton, Ohio: Ohio Psychology Press.

Stumpf, H. (1993). Performance factors and gender-related differences in spatial ability: Another assessment. *Memory & Cognition, 21*, 828-836.

Stumpf, H., & Eliot, J. (1995). Gender-related differences

in spatial ability and the *k* factor of general spatial ability in a population of academically talented students. *Personality and Individual Differences, 19,* 33-45.

Stumpf, H., & Jackson, D.N. (1994). Gender-related differences in cognitive abilities: Evidence from a medical school admissions testing program. *Personality and Individual Differences, 17,* 335-344.

Stumpf, H., & Klieme, E. (1989). Sex-related differences in spatial ability: More evidence for convergence. *Perceptual and Motor Skills, 69,* 915-921.

Stumpf, H., & Stanley, J. C. (1996). Gender-related differences on The College Board's Advanced Placement and Achievement tests, 1982-1992. *Journal of Educational Psychology, 88,* 353-364.

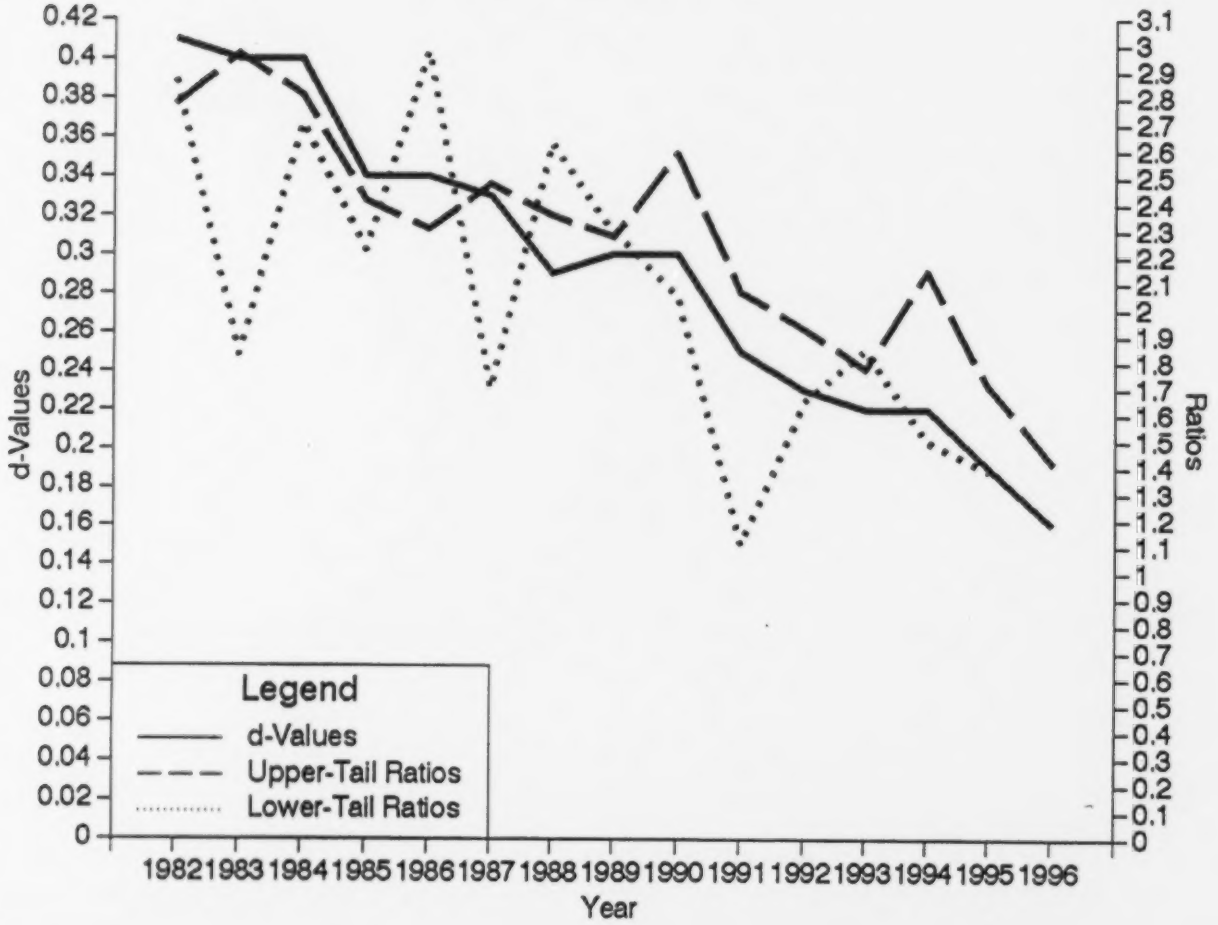
Stumpf, H., & Stanley, J. C. (1997). Gender-related differences in enrollment in and scores on the Advanced Placement Computer Science Examinations. Baltimore, MD 21218: The authors, IAAY, Johns Hopkins University.

Footnote

*Paper presented at the annual meeting of the Eastern Psychological Association in Washington, D. C., on 12 April 1997. Please address comments and inquiries to Professor Julian C. Stanley, SMPY, Bloomberg Center, Johns Hopkins University, Baltimore, MD 21218-2686, telephone (410) 516-6179, fax (410) 516-7239, e-mail setcty@jhu.edu .

Figure 1

Gender-Related Differences on the College Board Achievement Test in American History Over a Fifteen-Year Period



NOTE: No lower-tail ratio is reported for 1996, because in this year only four students attained the lower-tail score. The lower-tail ratios for 1982 through 1995 are based on 67 to 229 cases.

Figure 2

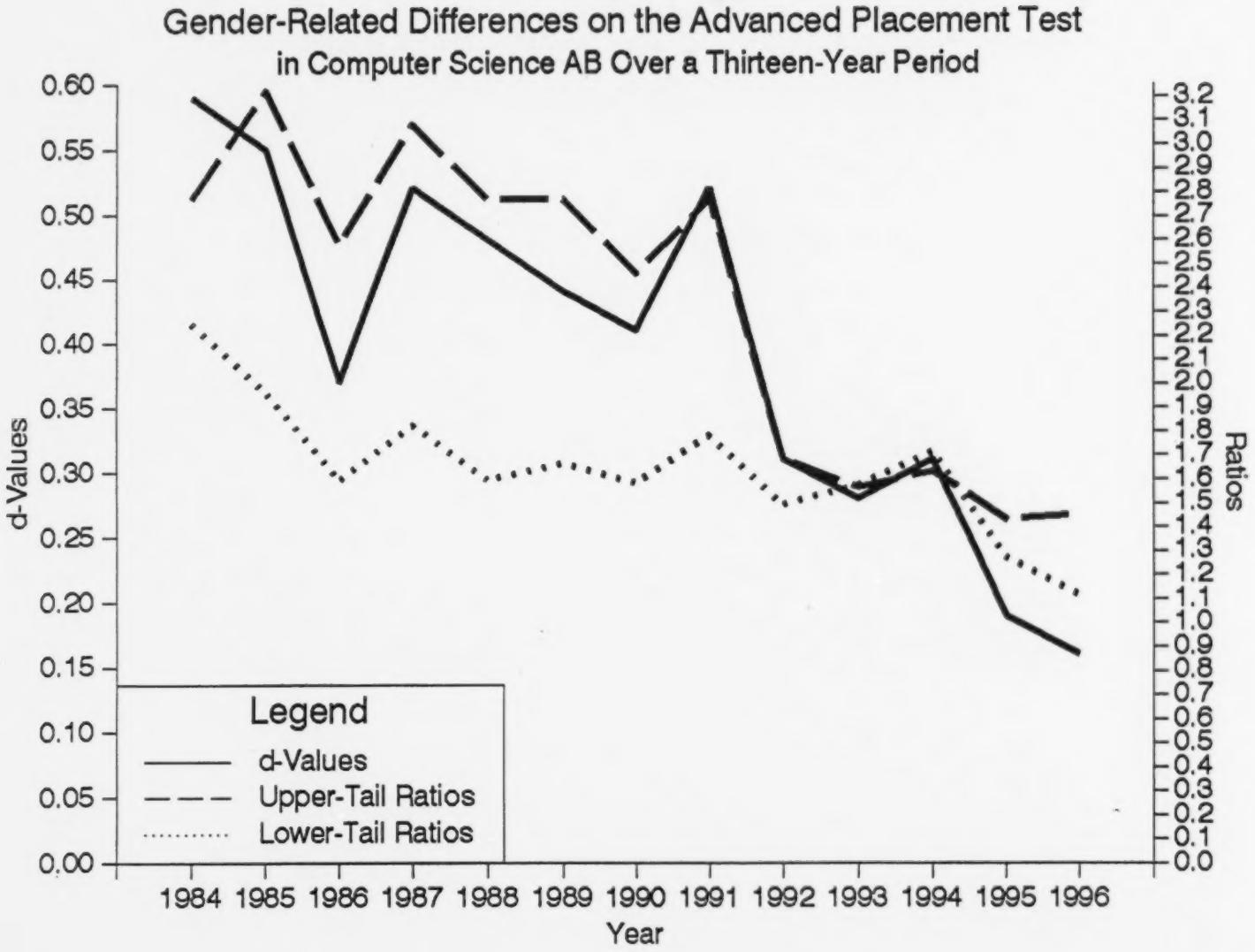

Gender-Related Differences on the Advanced Placement Test
in Computer Science AB Over a Thirteen-Year Period
d-Values
Ratios
Year
0.60
0.55
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
3.2
3.1
3.0
2.9
2.8
2.7
2.6
2.5
2.4
2.3
2.2
2.1
2.0
1.9
1.8
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996
Legend
d-Values
Upper-Tail Ratios
Lower-Tail Ratios

Figure 3

Gender-Related Differences on the Advancement Placement Test in Computer Science A Over a Six-Year Period

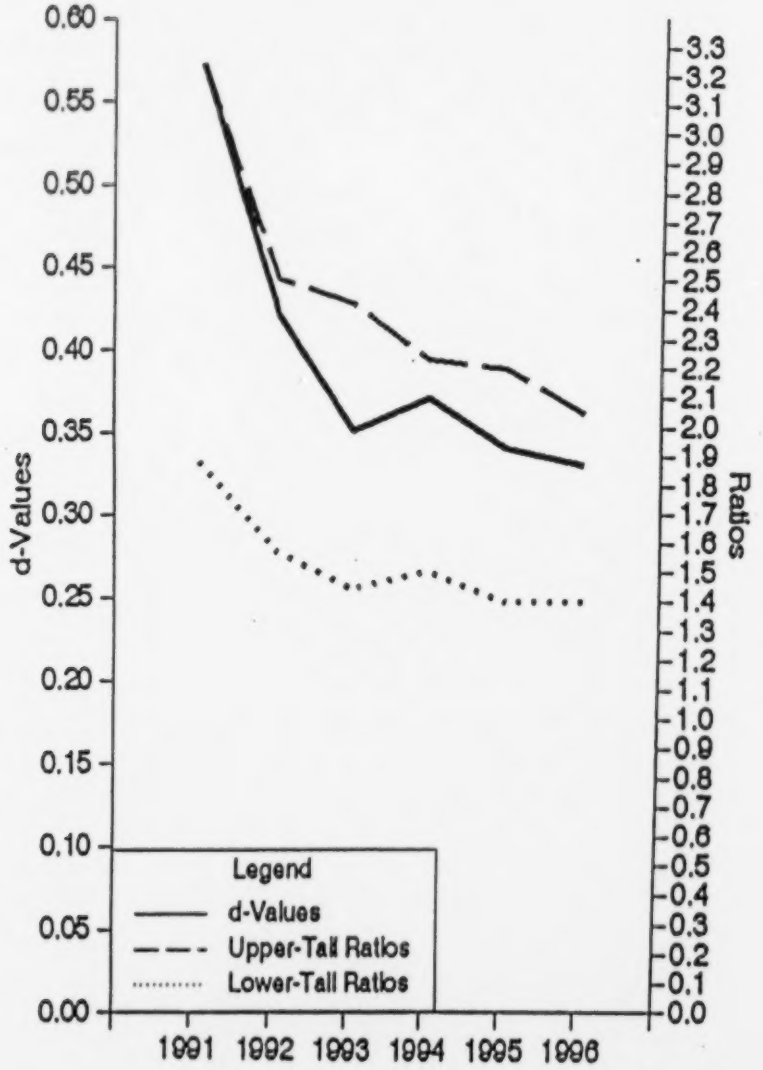


Table 1

Effect Sizes ≥ .20 and Upper-Tail Ratios ≥ 1.16 (Score ≥ 700) for the 21 College Board Achievement Tests (SAT II) Administered in 1996

Test	N	%Female	Effect Size	UTR	n on Which UTR is Based
12 FAVORING MALES:					
Physics	22,569	27	.50	2.14	5,815
Math Level II with Calculator	76,107	44	.42	1.86	21,511
World History	5,385	39	.37	2.40	820
Chemistry	41,215	44	.35	1.86	8,202
Math Level I	75,561	56	.32	2.05	7,680
Math Level I with Calculator	69,674	58	.29	1.79	6,166
Biology	52,909	55	.26	1.61	9,250
Latin	2,696	53	.23	1.44	523
American History	55,821	48		1.42	8,809
French	13,884	72		1.34	2,805
Modern Hebrew	848	58		1.34	166
Italian	616	66		1.28	200
2 FAVORING FEMALES:					
German with Listening	1,248	52		1.27	208
English Writing	198,381	54	(Lower-Tail Ratio: 1.85, n=276)		
7 FAVORING NEITHER SEX:	English Literature (N=45,103), Chinese with Listening (2,865), French with Listening (5,386), German (1,170), Japanese with Listening (1,379), Spanish (26,617), and Spanish with Listening (7,247)				
TOTAL: (FAVORING MALES)	225,221	54	.21	1.71	20,139

Table 2

Effect Sizes ≥ .20 and/or Tail Ratios ≥ 1.16 for the 29 Advanced Placement Program Examinations Administered in 1996

Test	N	%Female	Effect Size	UTR (Score = 5)	LTR (Score = 1)
18 FAVORING MALES:					
Physics C, Mechanics	11,072	26	.52	2.32	2.02
Physics B	18,664	35	.37	2.29	1.53
Economics, Macro	13,252	42	.37	1.76	1.91
Computer Science A	6,488	20	.33	2.05	1.40
Chemistry	37,462	42	.30	1.84	1.45
Government, U.S.	39,538	51	.29	1.73	1.67
Physics C, Elec. & Mag.	5,662	22	.28	1.59	1.51
Calculus BC	20,823	38	.27	1.50	1.41
Calculus AB	102,029	47	.26	1.76	1.37
Economics, Micro	10,025	40	.24	1.55	1.47
Biology	64,651	56	.24	1.39	1.54
History, U.S.	140,597	53	.23	1.46	1.56
Government, Comp.	5,781	45	.21	1.49	1.35
Computer Science AB	4,577	12		1.45	
European History	38,887	51		1.43	1.31
Art Studio General	5,901	58		1.20	*
Psychology	14,308	65		1.16	1.29
English Language	58,094	61		1.16	
6 FAVORING FEMALES:					
Spanish Literature	5,415	68	.23	1.37	1.87
French Literature	1,385	71	.20	1.20	1.65
Art Studio General	5,901	58			*1.36
German Language	2,941	53			1.32
English Literature	148,131	63			1.24
Art History	5,990	64			1.21
6 FAVORING NEITHER SEX:		Art Studio Drawing (N=2,635), French Language (11,987), Latin Literature (1,648), Latin Vergil (2,757), Music Theory (2,743), and Spanish Language (40,886)			
TOTAL (FAVORING MALES)	824,329	53		1.36	

Table 3

Intercorrelations of the Effect Sizes (*d*), Tail Ratios, and Percentages of Females Taking the 21 Achievement and 29 AP Tests

Achievement Tests

	d	UTR	LTR*
% Female	-.64	-.28	-.10
d		.86	.16
UTR			.23

*Few examinees scored less than 300, so these *r*s are unstable. Compare them with the *r*s below for the LTRs on the AP tests.

AP Tests

	d	UTR	LTR
% Female	-.71	-.62	-.66
d		.88	.86
UTR			.72